DOCUMENT RESUME

ED 228 295 TM 830 215

AUTHOR Rippey, Robert M.; And Others
TITLE The Validity of Computer Audits of Simulated Cases Records.
PUB DATE Apr 83
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Branching; Case Records; *Clinical Diagnosis; *Computer Assisted Testing; Computer Programs; Evaluation Methods; Feedback; *Medical Evaluation; Minicomputers; *Scoring; *Validity
IDENTIFIERS Computer Simulation; *Open Ended Questions

ABSTRACT

This paper describes the implementation of a computer-based approach to scoring open-ended problem lists constructed to evaluate student and practitioner clinical judgment from real or simulated records. Based on 62 previously administered and scored problem lists, the program was written in BASIC for a Heathkit H11A computer (equivalent to DEC PDP11/03). In addition to a scoring program, the complete package requires five other files: the scoring key, a dictionary of preferred expressions, the spelling algorithm, a program for preparing scoring keys, and a program for printing text files. The basic scoring structure consists of a short sequence of logical rules and a very long list of terms and equivalent expressions and stems which make up the logical data. A typical program output for one subject for one case would display not only counts of each error type, but also an accounting for each term on the problem list. The program provides a method of consistently evaluating scoreable problem lists without the use of faculty time. It also can be used for instructional purposes, not only providing a student with instantaneous feedback, but also with comments on performance in preparing a problem list from a set of clinical signs and symptoms. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

THE VALIDITY OF COMPUTER AUDITS OF SIMULATED CASES RECORDS

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Robert M. Rippey, Ph.D.
Anthony E. Voytovich, M.D.
Donald Jue, M.D.

University of Connecticut School of Medicine
Farmington, Connecticut 06032

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R.M. Rippey

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED228295

Clinical judgment is a complex behavior consisting of a variety of components. One important component is the ability to translate a set of signs and symptoms into a problem list. Successful manual methods of evaluating this aspect of student and practitioner performance have been developed, but the process of scoring is time consuming. This paper describes the implementation of a computer based approach to scoring open-ended student problem lists constructed from real or simulated records. The results are validated against expert human scorers.

Clinical reasoning has been studied globally by Elstein, Shulman, and Sprafka (1). Voytovich, Rippey, and Copertino (2) and Voytovich, Suffrendini, and Rippey (3) have studied a limited but important aspect of clinical judgment using simulated cases and scoreable problem lists. The most frequently encountered errors identified on scoreable problem lists are errors of inadequate specification of the disease entity, premature closures, listing of single clues as problems, and failure to identify problems which stand alone. These variables have been found to have high interrater reliability, clear scoring protocols which can be applied by non-physicians, generalizability of some measures across cases, and validity in separating subjects in terms of their level of training and experience. The errors have been found to apply to real case records as well as the simulated cases used in much of the research (4). In addition, the method has substantial educational value and has clarified the process of constructing problem lists for many students (5).

Over the past eight years, the method has evolved into a rather standard format as illustrated by Figure 1 of your handout. Student responses are collected on another standard form and a typical student response is shown in Figure 2. A key for each problem is constructed and may be revised as a result of encountering unusual student responses.

As Dr. Voytovich and I proceeded along the rocky road of validation, our enthusiasm waxed strong. Our conviction that we had a useful tool for medical educators never faltered. Our requests for reprints grew by leaps and bounds. However, we have not met many people who actually used this approach. We began to wonder, "Why not?"

Unfortunately, the scoring procedure is time consuming and raters needed some training. Indeed, after grading his 400th problem list one evening, even Tony began to complain. To compensate for the inconveniences and indignities concomitant with the grading process, we began to discuss

TM 830 215

Presented at the annual meeting of the American Educational Research Association, Montreal, April, 1983.

the possibility of a computer program to take the place of the trained rater. This computer program had the potential for not only replacing expensive manpower but also offering students rapid feedback for their problem lists when employed interactively. However, the task of programming seemed formidable. At this point Dr. Jue, a resident at the VA hospital in search of a summer research project and a skilled programmer, joined our group.

It was decided that, although a stylized list of problems would simplify scoring, a free form exercise would provide the most realistic picture of the student's ability to formulate a proper problem list de novo. On the other hand, the two perennial problems of open-ended responses to closed minded computers arose: spelling errors and unanticipated answers. An additional complication inherent to medical terminology was the large number of long words which were used frequently. To some extent, however, this turned out to be a blessing in disguise. Although long words require more extensive searching in a spelling algorithm, they occur less frequently than more common words, at least if one believes Zipf's law. In addition, the preciseness of reference of the long words means that there are fewer alternatives to confuse them with. Therefore, although the comparisons themselves may involve more letters than comparisons for short words, the number of words requiring comparison may be less. The alternatives to a misspelled word may be fewer.

METHODS

Sixty-two previously administered and scored problem lists, all related to the same case, were available in our files from a previous study of the reliability and validity of manually scored problem lists. The data had been obtained from a group of medical students, residents, and house staff at the University of Connecticut Health Center. Eighteen of these cases were selected at random to guide the initial development of the computer program. The scoring key and guidelines had been determined and the scores assigned by an expert physician five years prior to the present computer scoring (2). Scores of "complete specification," "stand alone problems," and "clues as problems" were initially selected for computer scoring although in the final program, all the original variables were scored as well refinements which emerged from our thinking about computer implementation. Thus the final set of variables which the computer detected included two positive attributes—fully specified complex problems and stand alone problems--in addition to seven errors: 1) inadequate specification or synthesis, 2) clues combined but not synthesized, 3) clues as problems, 4) premature closure, 5) simple and complex omissions, 6) wrong synthesis, and 7) totally wrong responses. Although previous studies have shown that the last two errors seldom occur, they were retained to account for the occasional anomaly.

An examination of the 18 initially selected student responses showed a variety of answer forms which were not readily classified by the manual scoring key and which required some limited judgment on the part of the scorer. Although a correct response may involve a phrase like "alcohol abuse," students can concoct a thousand ways of expressing it. On the other hand, it is not necessary to store up every possible phrase which students might emit, and the crucial skill in developing a computer scoring key is

recognizing what is important to identifying a correct response. The development of both a large set of synonyms and numerous proxes for relevant expressions solved most of the problems of variability of form of student expression. Proxes are reduced expressions which "stand in" for a variety of full expressions. The program itself operated by selecting or identifying any one acceptable set of conjunctions and disjunctions of a list of appropriate expressions and proxes. A portion of a scoring key is shown in Figure 3.

Spelling errors were identified by an algorithm which compared a standard vocabulary with each word in the problem list. Where matches were not found, three "nearest neighbors" were located and proposed as possible options. Where none of the proposed options was correct, the word was placed on an overflow list for further inspection. The original spelling program consumed an undue amount of time, taking about three minutes to identify words not in the standard vocabulary. A binary search modification was added which reduced the identification time to one minute or less which proved acceptable in practice. However, it was discovered that misspellings were not a formidable problem. In only a few cases did misspelled words alter the scoring. Therefore, the spelling algorithm was made optional. The resulting scoring time was reduced to less than 45 seconds per student. In practical terms this meant that the scoring was accomplished in about the same length of time it took to type in the problem list.

THE PROGRAM

The program was written in BASIC for a Heathkit H11A computer (equivalent to DEC PDP11/03). In addition to the scoring program, the complete package requires five other files: the scoring key which you have already seen, a dictionary of preferred expressions (Figure 4), the spelling algorithm, a program for preparing scoring keys, and a program for printing text files. The basic scoring structure comprises a short sequence of logical rules and a very long list of terms and equivalent expressions and stems which make up the logical data. One of the problem lists, for example, had to specify the complex problem, "pneumococcal pneumonia, left upper lobe." However, there are a large number of ways of saying this. Not only was the computer required to identify the conjunction of these five terms, but also to identify situations where the specification was incomplete, overly specified, or stated correctly in a different way. Some would say "upper left lobe." Others would say "pneumonia with pneumococci." To fill the answer key, the eighteen initial problem lists were useful. In addition, substantial conversation occurred among the authors anticipating the number of ways a particular problem could be stated. In all, about 560 words were used in the answer keys. These were sufficient to identify all but one word (septicemia) which occurred on the subsequently graded cases. A typical program output for one subject for one case would display not only counts of each error type, but also an accounting for each term on the problem list (Figure 5).

RESULTS

The 44 problem lists remaining after the program was developed were scored and a regression computed between the scores obtained by the human

scorer and the computer. If the human scorer and the computer were assigning equivalent scores, then the regression should have a slope of one and an intercept of zero. The results are shown in Table 1.

TABLE 1: PEARSON CORRELATIONS AND REGRESSIONS BETWEEN MAN AND MACHINE
DF = 42

Variable	Correlation	Regression (C=Computer, M=Man)
Clues as Problems	.979*	C = .96M - .05
Stand Alone	.859*	C = .80M + .04
Specification	.931*	C = 1.02M - .03

*p < .01

EDUCATIONAL IMPORTANCE

The program as developed provides a facility for consistently evaluating scoreable problem lists without the use of faculty time. It also can be used for instructional purposes, not only providing a student with instantaneous feedback, but also with comments on performance in preparing a problem list from a set of clinical signs and symptoms. This allows the student an opportunity to practice an important and sophisticated clinical behavior without the need of close faculty supervision. It also provides a standard approach to rapid scoring of an open-ended assessment of an important component of clinical judgment. The program is user cordial and readily amenable to an almost unlimited number of case problems. Copies of the program and keys are available from the first author.

REFERENCES

(1) Elstein, A., Shulman, L., and Sprafka, S. Medical Problem Solving: An Analysis of Clinical Reasoning, Cambridge, Massachusetts: Harvard University Press, 1978.

(2) Voytovich, A., Copertino, L. and Rippey, R. "Scoreable Problem Lists as Measures of Clinical Judgment," Evaluation & the Health Professions, 3(2):159-170, 1980.

(3) Voytovich, A., Suffrendini, A., and R. Rippey. "Overconfidence in Diagnosis," SREPCIM Abstracts, page 641A, Spring, 1981.

(4) Voytovich, A., Harper L., and Rippey, R. "Evaluating Analytic Reasoning in the Management of the Multiproblem Patient," Proceedings of the 18th Annual Conference on Research in Medical Education, 18:78-83, 1979.

(5) Voytovich, A. "Evolution of a Nonthreatening Audit in a Large Teaching Hospital," in Applying the Problem Oriented System, (H.K. Walker, J.W. Hurst, and M.F. Woody, eds.), New York: Medcom Press, 1973.

FIGURE 1
DATA BASE AND MANUAL SCORING KEY FOR A TYPICAL PROBLEM

A 60-year-old coal miner enters your office with a five-year history of increasingly severe dyspnea which has disabled him almost completely. He has four to five beers a day and drinks a quart of whiskey a week, mostly on weekends. He presents himself now because of a two to three day history of malaise, fever, chills, and a cough productive of large quantities of purulent and rusty sputum.

Physical findings include BP 170/100, p 120 and regular, temperature 102 orally. He appears terribly ill with a productive deep cough. Fundi show marked arterial narrowing and AV compression. There are rales, dullness to percussion, and egophony with increased tactile fremitis over the left upper lung field posteriorly. The liver is felt 5cm below the costal margin and is slightly tender. The remainder of the exam is unremarkable.

Hct = 40; WBC = 16,000 with 60 polys, 18 bands, 12 lymphs, 8 monos, 2 eos. Sputum is loaded with polys and gram positive diplococci. Chest film shows a dense lobar infiltrate (left upper lobe). A blood culture obtained the night before is growing Pneumococci.

Complex Problems

1. Pneumococcal pneumonia Left Upper Lobe (The addition of sepsis or positive blood culture is optional)

2. Hypertension with retinopathy

   or

   Hypertension with arteriolar narrowing. (A statement of "abnormal blood pressure" would constitute an underspecification)

Stand Alone Problems

3. Dyspnea
4. Alcohol use
5. Hepatomegaly
6. Occupational exposure to coal, silica (either/or) or just occupational exposure

PROBLEM LIST - Patient's Chart
HCH-179 REV. 9-75

UNIVERSITY OF CONNECTICUT HEALTH CENTER
JOHN DEMPSEY HOSPITAL
Farmington, CT 06032

PATIENT IDENTIFICATION ✓

NO.	ONSET DATE	ACTIVE PROBLEM		
		COPD (Emphysema) ↓ mobility to work		
		hazardous occupation (coal mining)		
		alcohol consumption (4-5 beers daily, 9 yrs of steady work)		
		active Pneumococcal infection 1) fever 2) productive cough		
		3) ↑ WBC (leukocytosis)		
		OR LUL infiltrate		
		enlarged liver		
		Atrial narrowing ≥ AV compression		
		hypertension 180/100		

BEST COPY AVAILABLE

```
PANS3.KEY
 PNEUMONIA
-1 , 0
 PNEUMOCOCCAL PNEUMOCOCCUS PNEUMOCOCCI STREP STREPTOCOCCUS
 1 , 0
 LUL LEFT L LOBAR
 1 , 0
 LUL UL UPPER LOBAR
 1 , 0
 LUL LUNG UL LOBAR
 1 , 0
 MALAISE ILL
-1 , 5
 HYPERTENSIVE HYPERTENSION HIGH INCREASED ELEVATED 170/100 ELEVATION
-2 , 0
 HYPERTENSIVE HYPERTENSION BLOOD BP
 2 , 0
 HYPERTENSIVE HYPERTENSION PRESSURE BP
 2 , 0
 RETINOPATHY RETINA RETINAL EYE EYES FUNDAL FUNDI FUNDOSCOPIC FUNDOSCO
PY K-W AV ARTERIAL ARTERIOLAR
 2 , 0
 RETINOPATHY CHANGES CHANGE ABNORMAL ABNORMALITY ABNORMALITIES K-W NAR
ROWING COMPRESSION NICKING
 2 , 0
 PHEOCHROMOCYTOMA
-2 , 6
 HYPERTENSIVE HYPERTENSION HIGH INCREASED ELEVATED ELEVATION 170/100
-2 , 4
 HYPERTENSIVE HYPERTENSION BLOOD BP HYPERTENSIVE HYPERTENSION BP PRES
URE
 2 , 4
HYPERTENSIVE HYPERTENSION BP PRESSURE
 2 , 4
 RETINOPATHY
-2 , 3
 RETINA RETINAL EYE EYES FUNDI FUNDAL FUNDOSCOPIC FUNDOSCOPY AV ARTERI
AL ARTERIOLAR
-2 , 3
 CHANGES CHANGE ABNORMAL ABNORMALITY ABNORMALITIES RETINOPATHY K-W NA
RROWING COMPRESSION NICKING
 2 , 3
 DYSPNEA SOB DOE SHORT SHORTNESS DIFFICULTY DIFFICULT DIFFICULTIES
-3 , 1
 DYSPNEA SOB DOE BREATH BREATHING
 3 , 1
 HEPATOMEGALY LIVER HEPAR HEPATIC HEPATOMEGALIA
-5 , 1
 HEPATIC CIRRHOSIS
-5 , 5
 OCCUPATION OCCUPATIONAL EXPOSURE COAL DUST SILICA MINER MINING ENVIRO
NMENT INDUSTRIAL
-6 , 1
 PNEUMOCONIOSIS SILICOSIS
-3 , 5
 BLACK
-3 , 6
 LUNG
 3 , 6
 C COPD CHRONIC
-3 , 6
 O COPD OBSTRUCTIVE
 3 , 6
 P L COPD PULMONARY LUNG
 3 , 6
 D COPD DISEASE
```

BEST COPY AVAILABLE

FIGURE 4: SHORT LIST OF PREFERRED EXPRESSIONS
AND LONG LIST OF A MUCH LONGER LIST OF STUDENT RESPONSES
(EVERYTHING THEY WANTED TO SAY ABOUT THE CASE AND DID NOT
HAVE THE DISCRETION TO SUPPRESS)

LUL PNEUMOCOCCAL PNEUMONIA
MALAISE

HYPERTENSION WITH RETINOPATHY
PHEOCHROMACYTOMA

HYPERTENSION
RETINOPATHY

RETINAL CHANGES

DYSPNEA
HEPATOMEGALY
LIVER DISEASE
OCCUPATIONAL EXPOSURE
PNEUMOCONIOSIS

BLACK LUNG

COPD
COUGH

FEVER
CHILLS

TACHYCARDIA
RALES
DULLNESS
EGOPHONY
FREMITUS
LEUKOCYTOSIS
LEUKOCYTOSIS

INFILTRATE
POSITIVE CULTURES
INFECTION/PNEUMONIA
ALCOHOL ABUSE
BACTEREMIA
SHOCK

PULMONARY EMBOLUS
CARDIORESPIRATORY DISEASES
CHF

DYSPNEA
ECCYHYMOSIS
EDEMA
EGOPHONY
ELEVATED
ELEVATION
EMBOLI
EMBOLISM
EMBOLUS
EMPHYSEMA
ENDOCARDITIS
ENTERS
ENVIRONMENT
EOS
EPIDERMOID
EPISTAXIS
EQOPHONY
ERYTHROBLAST
ERYTHROCYTE
ERYTHROPOIESIS
ETOH
ETOHISM
EXAM
EXPOSURE
EXTENSION
EXUDATE
EYE
EYES
FAILURE
FAST
FAT
FELT
FEVER
FIBROMA
FIBROSIS
FIELD
FILM
FINDINGS
FIVE-YEAR
FOCAL
FREMITUS
FUNDAL
FUNDI
FUNDOSCOPIC
FUNDOSCOPY
FUNDSCOPY
FUNGUS

GANGRENE
GLOMERULONEPHRITIS
GLOMERULOSCLEROSIS
GRAM
GRANULOCYTE
GRANULOMA
GROWING
HAMARTOMA
HAS
HCT=40
HE
HEART
HEMATEMESIS
HEMATOMA
HEMATURIA
HEMOCHROMATOSIS
HEMOPTYSIS
HEMORRHAGE
HEMOSIDERIN

FIGURE 5: EXAMPLE OF STUDENT RESPONSE
AND COMPUTER SCORING IN INTERACTIVE MODE
(NOTE THE TWO SPELLING ERRORS)

```
?HAZARDOUS OCCUPATION COAL MINING
HAZARDOUS OCCUPATION COAL MINING
HAZARDOUS NOT UNDERSTOOD
PRESS <RETURN> IF CORRECT AS IS OR
RETYPE IF MISPELLED OR 'S' TO TRY SPELLING PROGRAM
?
O.K.
?O
**********************************************************************
 OCCUPATIONAL EXPOSURE=SIMPLE PROBL


?ALCOHOL CONSUMPTION 4 5 BEERS DAILY QUART OF WHISKEY WEEKLY
ALCOHOL CONSUMPTION 4 5 BEERS DAILY QUART OF WHISKEY WEEKLY
CONSUMPTION NOT UNDERSTOOD
PRESS <RETURN> IF CORRECT AS IS OR
RETYPE IF MISPELLED OR 'S' TO TRY SPELLING PROGRAM
?O
O.K.
DAILY NOT UNDERSTOOD
PRESS <RETURN> IF CORRECT AS IS OR
RETYPE IF MISPELLED OR 'S' TO TRY SPELLING PROGRAM
?
O.K.
WEEKLY NOT UNDERSTOOD
PRESS <RETURN> IF CORRECT AS IS OR
RETYPE IF MISPELLED OR 'S' TO TRY SPELLING PROGRAM
?
O.K.
?O
**********************************************************************
 ALCOHOL ABUSE=SIMPLE PROBL


?ACTIVE PNEUMOCOCCI INFECTION FEVER PRODUCTIVE COUGH ELEVATED WBC LEUKOCYTOSIS
ACTIVE PNEUMOCOCCI INFECTION FEVER PRODUCTIVE COUGH ELEVATED WBC LEUKOCY
TOSIS
ACTIVE NOT UNDERSTOOD
PRESS <RETURN> IF CORRECT AS IS OR
RETYPE IF MISPELLED OR 'S' TO TRY SPELLING PROGRAM
?
O.K.
?O
**********************************************************************
( INFECTION/PNEUMONIA =INADEQ. SPEC.)


?LUL INFILTRATE
LUL INFILTRATE
?O
**********************************************************************
( INFILTRATE =CLUE AS PROBLEM)


?ENLARGED LIVER
ENLARGED LIVER
ENLARGED NOT UNDERSTOOD
PRESS <RETURN> IF CORRECT AS IS OR
RETYPE IF MISPELLED OR 'S' TO TRY SPELLING PROGRAM
?
O.K.
?O
```

BEST COPY AVAILABLE

FIGURE 5 (CONTINUED)

```
**********************************************************************
 RETRIEVAL=SIMPLE PROBL


 CRITERIA NARROWING AND AV COMPREHENSION
 ARTERYAL NARROWING AND AV COMPREHENSION
 ARTERYAL NOT UNDERSTOOD
 PRESS <RETURN> IF CORRECT AS IS OR
 RETYPE IF MISPELLED OR 'S' TO TRY SPELLING PROGRAM
 ?S
 DID YOU MEAN ARTERIAL OR ARTERIOLAR OR ARE OR SOMETHING ELSE
 ENTER WORD
 ?ARTERIAL
 O.K.
 COMPREHENSION NOT UNDERSTOOD
 PRESS <RETURN> IF CORRECT AS IS OR
 RETYPE IF MISPELLED OR 'S' TO TRY SPELLING PROGRAM
 ?S
 DID YOU MEAN COMPRESSION OR COMPLEX OR COMPLETELY OR SOMETHING ELSE
 ENTER WORD
 ?COMPRESSION
 O.K.
 ?0
**********************************************************************
 ( RETRIEVE CHANGES =CLUE AS PROBLEM)


 ?HYPERTENSION 180/100
 HYPERTENSION 180/100
 180/100 NOT UNDERSTOOD
 PRESS <RETURN> IF CORRECT AS IS OR
 RETYPE IF MISPELLED OR 'S' TO TRY SPELLING PROGRAM
 ?
 O.K.
 ?000
**********EXTENSION**********************************************************
 ( HYPERTENSION =INADEQ. SPEC.)


 OMISSION OF PROBLEM/CLUE OF  3


 STUDENT
 FULLY SPECIFIED COMPLEX PROBLEMS = 0
 STAND ALONE PROBLEMS             = 3

 ERRORS
 INADEQUATE SPEC.  2
 CLUES COMBINED/EXT INADEQUATE SYNTHESIS 0
 CLUE AS PROBLEM   2
 PREMATURE CLOSURE 0
 OMISSIONS' SIMPLE 1
            COMPLX 0
```

BEST COPY AVAILABLE